DOCUMENT RESUME

ED 204 404 TM 810 418

AUTHOR Benson, Jeri: Wilcox, Stuart
TITLE The Effect of Positive and Negative Item Phrasing on the Measurement of Attitudes.
PUB DATE Apr 81
NOTE 13p.: Paper presented at the Annual Meeting of the National Council on Measurement in Education (Los Angeles, CA, April 11-17, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Age Differences: Intermediate Grades: Racial Integration: *Response Style (Tests): Sex Differences: *Student Attitudes: *Test Construction: *Test Format: Test Items

ABSTRACT

What effect varying the phrasing of attitude statements would have across sex and grade was determined. Three content parallel forms were developed to assess student attitudes toward integration. Each form contained 15 items and differed only in terms of item phrasing either positive, negative or a combination of both. Six hundred and twenty-two, fourth, fifth, and sixth grade students responded anonymously to one of the three forms. The data indicated that positive, negative, and mixed item phrasing resulted in differential response patterns across grade level. In addition, differences were noted in the internal consistency estimates across forms. Implications of the study were discussed from a test constructor-test user perspective. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED204404

The Effect of Positive and Negative Item Phrasing on the Measurement of Attitudes

Jeri Benson & Stuart Wilcox

University of Southern California

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Benson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810 418

A paper presented at the annual meeting of the National Council on Measurement in Education, Los Angeles, California, April, 1981.

The Effect of Positive and Negative Item Phrasing
on the Measurement of Attitudes

ABSTRACT

The purpose of the study was to determine what effect varying the phrasing of attitude statements would have across sex and grade. Three content parallel forms were developed to assess student attitudes toward integration. Each form contained 15 items and differed only in terms of item phrasing either positive, negative or a combination of both. Six hundred and twenty-two, fourth, fifth, and sixth grade students responded anonymously to one of the three forms. The data indicated that positive, negative, and mixed item phrasing resulted in differential response patterns across grade level. In addition, differences were noted in the internal consistency estimates across forms. Implications of the study were discussed from a test constructor-test user perspective.

The Effect of Positive and Negative Item Phrasing on the Measurement of Attitudes

The development of items to assess student attitudes poses very complex methodological problems for the test constructor such as a definition of the attitude, reliability and validity concerns, response sets, and more recently, issues of bias. One of the more pervasive of these problems has been the question of whether a response set may have affected the results and hence the interpretation of an attitude survey. The possibility of a response set influencing the validity of cognitive instruments was first reported by Cronbach (1946; 1950). He defined response set as "any tendency causing a person consistently to make different responses to test items than he would have had the same content been presented in a different form" (1946, p. 476). Rorer (1965) after reviewing the literature on response sets in cognitive and noncognitive instruments concluded an individuals response set can be influenced by item content and may account for a significant proportion of response variance. Mehrens and Lehmann (1980) have cited numerous research articles that have attempted to resolve many of the methodological problems of response sets and noncognitive assessment; however, no diffinitive solution has yet been found.

The importance of reducing the effects of a response set in school evaluation research, where cognitive and noncognitive data are to be gathered, is crucial if one is to collect valid data on

which to base evaluation conclusions and recommendations. Several procedures have been developed to control or limit the operation of response sets in cognitive instruments, e.g. directions to reduce guessing, writing clear concise items, randomly positioning the correct answer etc. (Guilford, 1967). In order to reduce the possibility of a response set operating in noncognitive instruments, it has been suggested to use a forced choice format, have an equal number of positive to negative items, clearly defining the response format etc. (Mehrens & Lehmann, 1980). Since evaluators are increasingly using affective instruments to gather data on the effects of curricular programs that impact on students, it is encumbent upon psychometricians to carefully study how the above suggestions may or may not be an aid to reducing response set and hence increasing the reliability and validity of data interpretations.

Recent research in the area of constructing attitude items (DeGracie & Vivino, 1977; Frisbie & Bradenburg, 1979) has indicated that the type of response format may alter an examinee's response. If varying the response format will elicit different responses from examinees to the same attitude statement, what effect will varying the phrasing of the attitude item have on an examinee's response? Most attitude scales are developed to include both positively and negatively phrased items in an attempt to reduce the possibility of a response set occuring (Cronbach, 1950; Coach & Keniston, 1960). This technique is used regardless of the age of the examinee. The purpose of the present study was to empirically compare the effects of varying the phrasing on attitude statements, using either all

positive, all negative, or a balance of positive and negative items, on an elementary school population.

METHOD

Sample

The sample consisted of all fourth, fifth, and sixth grade students (N=622) from three schools who were part of a district-wide court-ordered integration plan. Seventy-eight of the cases had to be dropped from the analyses due to a lack of identifiable information regarding the sex of the student, thus, 544 student responses were analyzed. The demographic breakdown is shown in Table 1.

Put Table 1 about here

Since the data were collected as a part of an assessment of an integration study at the three schools, the researchers were not allowed to ask the respondents their ethnic origin. However, the enrollment figures at the three schools indicated that approximately 30 percent of the students were White, 28 percent were Hispanic, and 42 percent were Black.

Instrumentation

Three content parallel forms were developed to assess student attitudes toward integration. The forms differed only in terms of item phrasing. The first form contained 15 positively phrased items. The second form contained the same 15 items negatively phrased. The third form contained the same 15 items, eight were positively phrased and seven were negatively phrased. All items

were presented in the same sequence.

Items written for the three forms were developed based upon objectives the schools proposed during the first year of court-ordered integration. Thus, a student's attitude toward integration was defined by the responses to items developed to assess the school's objectives. The items were judged for content relevence by the persons responsible for conducting the integration program at each school. An example of an item from two of the forms is presented below:

Positive Form- I like going to an integrated school.

Negative Form- I do <u>not</u> like going to an integrated school.

The items appearing on the Mixed Form were exactly like items on the Positive and the Negative Forms with the exception that eight were positively phrased and seven were negatively phrased.

Procedures

Each student responded to one of the three forms which were randomly administered in each class to all students by their classroom teacher at the end of the Spring semester, 1980. The students responded anonymously to the items on the attitude surveys by marking separate machine scorable answer sheets.

Each item was scored by weighting the responses on a 1-5 scale (1=strongly agree, 5=strongly disagree). A three way analysis of variance (ANOVA) was run on the dependent variable of student attitude. The three independent variables were sex, form, and grade. Alpha was set at .05. Since the school's would not allow the researchers to ask the ethnic orgin of each student, that variable was not a part of the analysis.

Results and Discussion

The responses from the student surveys were analyzed to a) obtain reliability estimates for the three forms, and b) obtain statistical analyses on the response differences, if any, for the three forms across the independent variables of sex, grade, and form.

Coefficient alpha was calculated to obtain a measure of each form's internal consistency. After reverse keying the negatively phrased items on the Mixed and Negative Forms, the following coefficients were obtained.

Positive Form	r_{xx}=.78	N=212
Negative Form	r_{xx}=.78	N=201
Mixed Form	r_{xx}=.63	N=203

The reliability of the Positive Form and the Negative Form was quite adequate for a 15 item attitude survey. However, the reliability of the Mixed Form, where there were eight positively phrased items and seven negatively phrased items, was markedly lower than the all Positive and all Negative Form. Since the content of the items for the Mixed Form was identical to that of the Positive and Negative Forms, the lower reliability of the Mixed Form may be an indication that the elementary students were having difficulty with the form and as such responded inconsistently to items where the phrasing was mixed.

Following the reliability analyses, a 2x3x3 ANOVA was run on the student responses where there were two levels for sex, and three levels for grade and form. The results indicated that the three way interaction was not significant ($F(4,526)= 1.34$, $p > .05$). Since the main effect for sex and the two-two way

interactions involving sex were nonsignificant, the variance due to the three-way interaction was collapsed and a 3x3 ANOVA was run using only grade and form.

The results of the grade by form ANOVA indicated that the two-way interaction again was statistically significant ($F(4,591)=3.54$, $p < .05$), and the two main effects for grade and form were both significant ($F(2,591)=6.94$, $p < .05$ and $F(2,591)=3.99$, $p < .05$ respectively). By dropping the variable sex, additional cases that were previously not identifable, were available for the two-way ANOVA (N=600). A simple main effects analysis was performed on the grade by form interaction to find at which level the significant difference was occuring. The data indicated that form at grade four was significant ($F(2,591)=6.93$, $p < .05$), grade for the Mixed Form was significant ($F(2,591)=6.96$, $p < .05$), and grade for the Negative Form was significant ($F(2,591)=6.74$, $p < .05$). Post hoc comparisons were run for each of the three significant findinds using Tukey's HSD procedure with alpha set at .05. The means and standard deviations across grade level and forms are reported in Table 2.

Put Table 2 about here

The results are summarized and discussed below. Recall that the strongly agree response was coded one and the strongly disagree response was coded five, therefore the lower the mean score the more favorable the attitude.

- The fourth grade students responded more favorably toward integration on the Positive Form than on the Negative or Mixed Forms ($\bar{X} = 35.44$, $\bar{X} = 39.43$, $\bar{X} = 40.66$ respectively)

- The sixth grade students responded more favorable toward integration on the negatively phrased form than did either the fifth or fourth grade students ($\overline{X}$ = 34.95, $\overline{X}$ = 40.66, $\overline{X}$ = 38.57 respectively)
- The sixth grade students responded more favorably toward integration on the mixed form than the fourth grade students ($\overline{X}$ = 33.79, $\overline{X}$ = 39.43 respectively)

Based upon the above findings, it was concluded that responses to an attitude questionnaire were significantly related to item phrasing and student grade level. In terms of item phrasing, younger students (fourth graders) tended to mark more favorable responses when the items were all positively phrased than when the same items were either all negatively phrased or, when a mix of positively and negatively phrased items was presented. In contrast, varying the item phrasing did not appear to be a problem for the older students (sixth graders) as their responses were not significantly different across the three forms. In addition, the older students tended to be able to handle the negatively phrased items as well as a mix of positive and negative items better than the younger fourth and fifth grade students.

Conclusion

Educators need to be aware of the possible effects item phrasing may have on the assessment of attitudes in children. In terms of controlling response sets in noncognitive instruments, the recomendation to use an equal number of positively and negatively phrased items may do more harm to a research study or evaluation than it was

originally thought to correct, especially when dealing with younger students. In terms of the reliability of the attitude measured, mixing the item phrasing may elicit inconsistent responses from the younger examinee, thus, resulting in lower reliability. In terms of the validity of the attitude measured, is one really improving validity by using positive and negative item phrasing? Although the construct validity of the three forms in the present study was not assessed, previous researchers have found that positive and negatively phrased items did not load on the same factor (Campbell & Grissom, 1979) and that positive and negative item structure significantly influenced student responses (Simpson, Rentz, & Shrum, 1976). Thus, a balance of positive and negatively phrased items may affect the reliability and validity of a test score in ways that are not intended. Additional research is needed on the effect item phrasing has on the psychometric properties of noncognitive measures for various populations. In the mean time, item phrasing should be carefully considered prior to the development or use of a noncognitive instrument with young children.

REFERENCES

Campbell, N. & Grissom, S. Influence of item direction on student responses in attitude measurement. 1979. (ERIC Document #ED 170 366)

Couch, A. & Keniston, K. Yeasayers and Nayseyers: Agreeing response set as a personality variable. Journal of Abnormal and Social Psychology, 1960, 60, 151-174.

Cronbach, L. J. Further evidence on response sets and test design. Educational and Psychological Measurement, 1950, 10, 30-31.

Cronbach, L. J. Response sets and test validity. Educational and Psychological Measurement, 1946, 6, 616-623.

DeGracie, J. S. & Vicino, F. L. Yes-No and all of the time to never. Educational Technology, 1977, 17, 34-36

Frisbie, D. & Brandenburg, D. Equivalence of questionnaire items with varying response formats. Journal of Educational Measurement, 1979, 16, 43-48.

Guilford, J. P. Response biases and response sets. In M. Fishbein (ed.) Readings in Attitude Theory and Measurement, New York, N.Y.: Wiley, 1967.

Mehrens, W. & Lehmann, I. Standardized tests in education (3rd ed.) New York, N.Y.: Holt Rinehart & Winston, 1980.

Rorer, L. G. The great response style myth. Psychological Bulletin, 1965, 63, 129-156.

Simpson, R. D., Rentz, R. R., & Shrum, J. W. Influence of instrument characteristics on student responses in attitude assessment. Journal of Research in Science Teaching, 1976, 13, 275-281.

Table 1

Demographic Characteristics of the Sample

Sex	Grade 4	5	6	Totals
Male	94	97	81	266
Female	95	98	85	278
Totals	189	189	166	544

Table 2

Means and Standard Deviations Across Grade Level and Form

Grade	Form: Positive	Mixed	Negative	Overall
4th	35.44 (7.64)	39.43 (6.94)	40.66 (10.66)	38.41 (8.70)
5th	36.09 (9.22)	35.95 (7.29)	38.51 (9.50)	36.81 (8.74)
6th	36.74 (7.90)	33.79 (7.07)	34.95 (9.00)	35.23 (8.10)
Overall	36.06 (8.29)	36.63 (7.43)	38.13 (9.97)	

Note: The responses were keyed such that the lower score represents a more favorable attitude.